STATISTICS 244 **Problem Set 5** **9:30-10:50 TTh**

Due Thursday February 21, 2019

1. Many non-normal distributions are approximately normal, $N(\mu,\sigma^2)$. One example is the Beta(α,β) distribution, as long as α and β are moderately large and the Beta expectation and variance are used for μ and σ^2. Suppose X has a Beta density. (i) Use the normal approximation to find all three of $P(|X-.6|<.001)$, $P(|X-.6|<.01)$, and $P(|X-.6|<.1)$, when X is (a) Beta(3,2), (b) Beta(30,20), (c) Beta(300, 200). (Present the results in a 3x3 table.) (ii) For case (a) only, find the same three probabilities exactly directly from the beta density by integration.

2. Election ties. What is the chance that there will be an exact tie in an election? Suppose there are n voters and that n is even (otherwise a tie is impossible). Suppose the voters vote independently of one another, each with probability p of voting for candidate A and probability 1 – p of voting for candidate B. Let X = total vote for A; then there will be a tie only if X =n/2.
(a) What is the probability of a tie if p = .5 and n = 20?
(b) What is the probability of a tie if p = .6 and n = 20?
(c) Suppose p is unknown, and our uncertainty about it is described adequately by a Beta (4,4). What is the probability of a tie (i.e. the marginal probability P(X = n/2)), if n = 20?
[Note: The comparison here has a bearing on a question in political science, namely why do people vote? Some might argue they vote because they want to break a potential tie. But that is a remote possibility under this model. Which of these models is most favorable to the theory? Least favorable? These calculations are only an example, but using Stirling's Formula you can find approximate probabilities for any n, for all these cases.]

3. Based on student A's performance during the first two weeks of a course, the professor has approximately a normal $N(75,8^2)$ prior distribution about the student's true ability, on a scale of 0 to 100. Consider the midterm examination as an error-prone measure of the student's true ability, where if the true ability is x, the examination score can be modeled as approximately normally distributed, $N(x, 6^2)$. The student scores 90 on the midterm; what are the posterior expectation and the probability that the student's true ability is above 85? Above 90?

4. Suppose $\hat{\theta}_1$ and $\hat{\theta}_2$ are uncorrelated and both are unbiased estimators of θ, and that $\mathrm{Var}(\hat{\theta}_1) = 2\mathrm{Var}(\hat{\theta}_2)$. (a) Show that for any constant c, the weighted average $\hat{\theta}_3 = c\hat{\theta}_1 + (1-c)\hat{\theta}_2$ is an unbiased estimator of θ. (b) Find the c for which $\hat{\theta}_3$ has the smallest MSE. (c) Are there any values of c ($0 \le c \le 1$) for which $\hat{\theta}_3$ is better (in the sense of MSE) than both $\hat{\theta}_1$ and $\hat{\theta}_2$? Which values?

5. Suppose the data consist of a single number X, and the model is that X has the following probability density:

$$f(x\,|\,\theta) = (1 + x\theta)/2 \quad \text{for} \quad -1 \le x \le 1; \quad = 0 \text{ otherwise.}$$

Supposing the possible values of θ are $0 \le \theta \le 1$; find the maximum likelihood estimate (MLE) of θ, $\hat{\theta}$, and find its (exact) probability distribution. Is the MLE unbiased? Find its bias and MSE. [Hint: First find the MLE for a few sample values of X, such as X = –.5 and X = .5; that should suggest to you the general solution. Drawing a graph helps! The distribution of the MLE will of course depend upon θ.]

6. In a famous example, Bortkiewicz tabulated the counts of deaths by horsekick in the Prussian Cavalry, for 14 Corps over 20 years (1875-1894), giving n = 280 observations in all. The frequency tabulation was:

Number of deaths	Frequency count
0	144
1	91
2	32
3	11
4	2
More	0
Total	280

These represent data $X_1, \ldots, X_n$, and one model that has been considered for these data is the Poisson distribution:

$P(X_i = k \mid \theta) = e^{-\theta}\theta^k / k!$, for k = 0, 1, 2, ...

The parameter θ equals both the expectation $E(X_i)$ and variance $V(X_i)$ for this distribution. It is thus the mean number of deaths per year for a single Corps.

Solve (a) – (d) algebraically (symbolically), then use the data to answer (e).

(a) For this model find the MLE of θ, assuming the X_i's are independent.
(b) Find the MSE of the MLE.
(c) From the Central Limit Theorem find the approximate distribution of the MLE when n is large.
(d) From Fisher's Approximation, find the approximate distribution of the MLE when n is large.
(e) Evaluate the MLE for the given data.
(f) If θ, the mean number of deaths per Corps in a year, is really 1.0, what (approximately) is the probability that the MLE would turn out to be below 0.85?

Note on the Poisson Distribution. If T has a Poisson distribution, Pois(θ), its probability function is given by $p(t \mid \theta) = \dfrac{\theta^t e^{-\theta}}{t!}$, for $t = 0, 1, 2, \cdots$.

Then E(T)= θ and Var(T) = θ. If X and Y are independent, each with a Poisson distribution, then X+Y has a Poison distribution. If X has a Binomial (n,p) distribution, and if p is small, X has approximately a Pois(np) distribution: $b(k;n,p) \approx p(k \mid np)$.

7. Suppose that X is the number of successes in a Binomial experiment with n trials and probability of success θ/(1+θ), where $0 \leq \theta < \infty$. (a) Find the MLE of θ. (b) Use Fisher's Theorem to find the approximate distribution of the MLE when n is large.